

Architecting High Availability Lustre Storage Solution - ClusterStor 6000

John Fragalla

Principal Solutions Architect
High Performance Computing

xyratex.

HPC Storage Trends

- High throughput storage is expected to be reliable & available
- Downtime is not tolerable
- High Availability & Reliability are becoming standard features
 - Increasing trend, ranked higher than throughput & capacity
- Focus on storage in many HPC centers to increase overall performance for applications
 - Adding compute without addressing storage does not improve application performance

Basis for Building Reliable HPC Storage

- It starts with the disk drives
- Redundancy in the hardware
- Software built with high availability features
- End to end software management
- Integration & testing



Disk Drive Testing and Integration

- First Phase
 - Rigorous disk drive testing before integrating the solution into a chassis
- Second Phase
 - Disk drive testing integrated into a chassis
 - Running various low level testing for connectivity
 - Vibration testing
- Third Phase
 - Disk drive testing with the software installed
 - Rigorous test cases

Hardware Redundancy

- No single point of failure for the entire HPC data storage solution
- Phase 1 – Power
 - Redundant power supplies, fans, & rack PDUs
 - Multiple power grids
 - UPS
- Phase 2 – Servers
 - Configure Lustre[®] servers in HA pairs
- Phase 3 – Management
 - Redundant end-to-end management from bare metal to filesystem
- Phase 4 – High Speed Network
 - Redundant fabric switches
 - Tolerate a complete switch failure

Software Resiliency and Reliability

- Software is based high availability to configure, manage, and monitor services
- Tolerate any type of failure to provide continuous data access
- Other benefits of High Availability
 - Live software upgrades
 - Live system maintenance
 - Live hardware upgrades/maintenance
- Data protection layer
 - No need to utilize dedicated RAID controllers

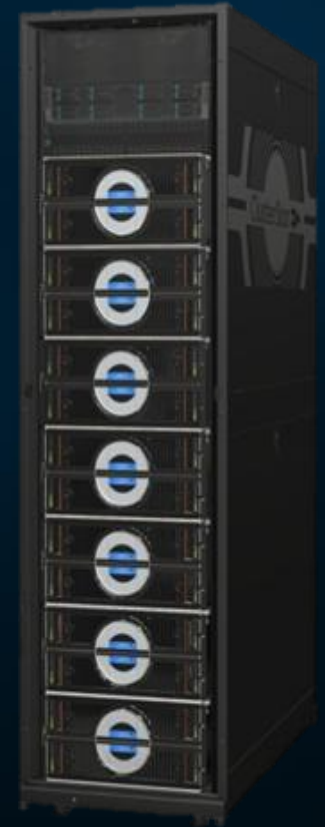
Integration and Solution Testing in Factory

- Hardware integration
 - Integrating the storage solution designed from the ground up into a single chassis proves to be highly available compared to non-integrated solutions
- Building, configuring and testing the entire storage solution in factory improves reliability and reduces on-site integration
- Repeatable performance benchmarks for proven throughput

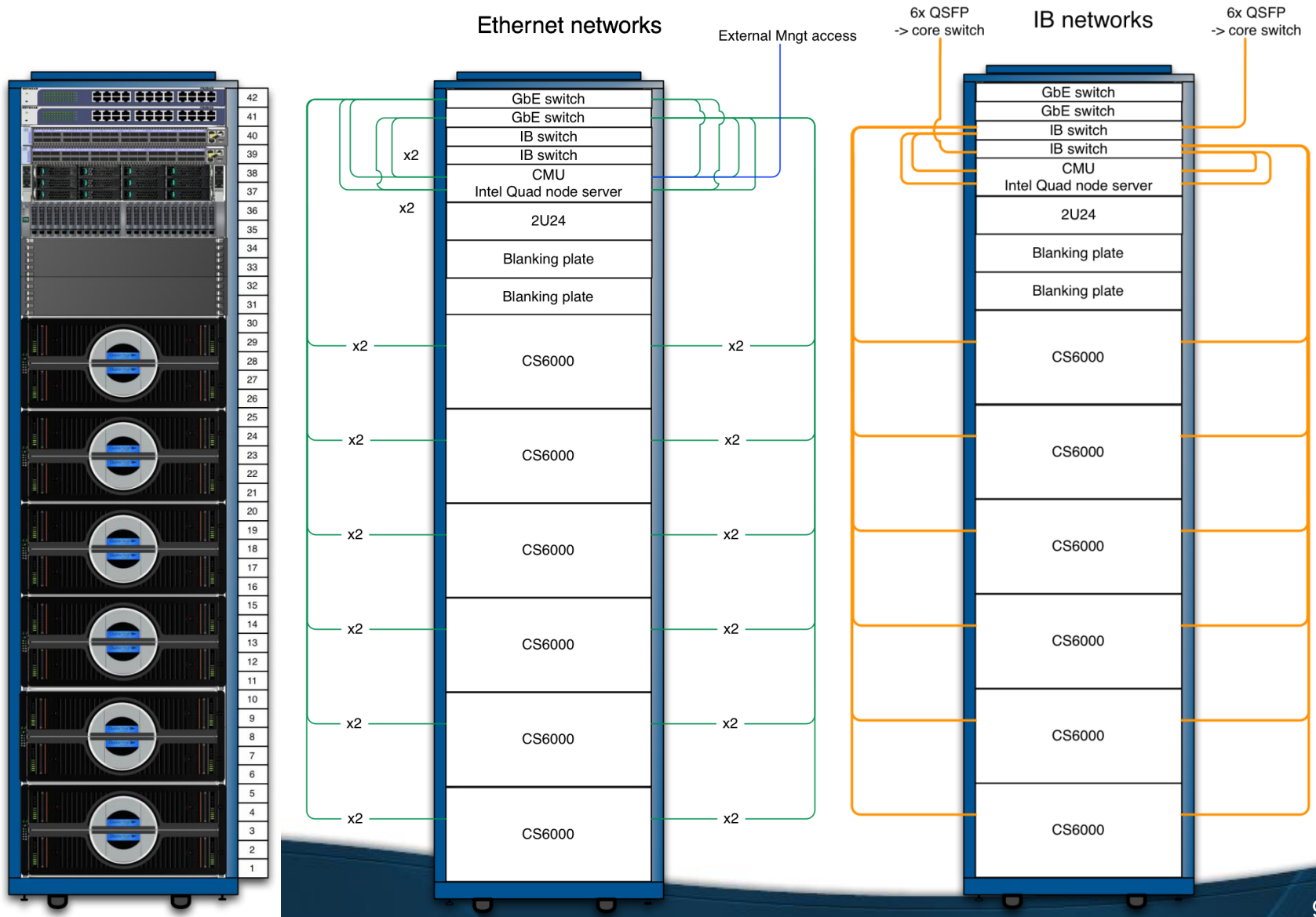
ClusterStor™ 6000 Overview

- Up to 42 GB/s per rack of Lustre filesystem performance
- Overall Performance scalable to > 1TB/s bandwidth
- Overall capacity scalable to > 100PBs
- Support FDR or 40GigE
- ClusterStor a complete ready-to-run Lustre solution
 - Up to 560 Data HDD's per rack (42RU)
 - Up to 1.8PBs usable per rack (with 4TB HDD's)
 - Up to 14 Lustre OSS servers per rack
 - Up to 14 high bandwidth network connectivity ports/rack
 - Factory Integration & staging
 - Rack integration & cabling
 - Entire storage software stack factory pre-installed & pre-configured
 - System Burn-in & benchmark testing area at Xyratex factory
 - “Rack'n'Roll” installation – hours vs. days or weeks

ClusterStor™ 

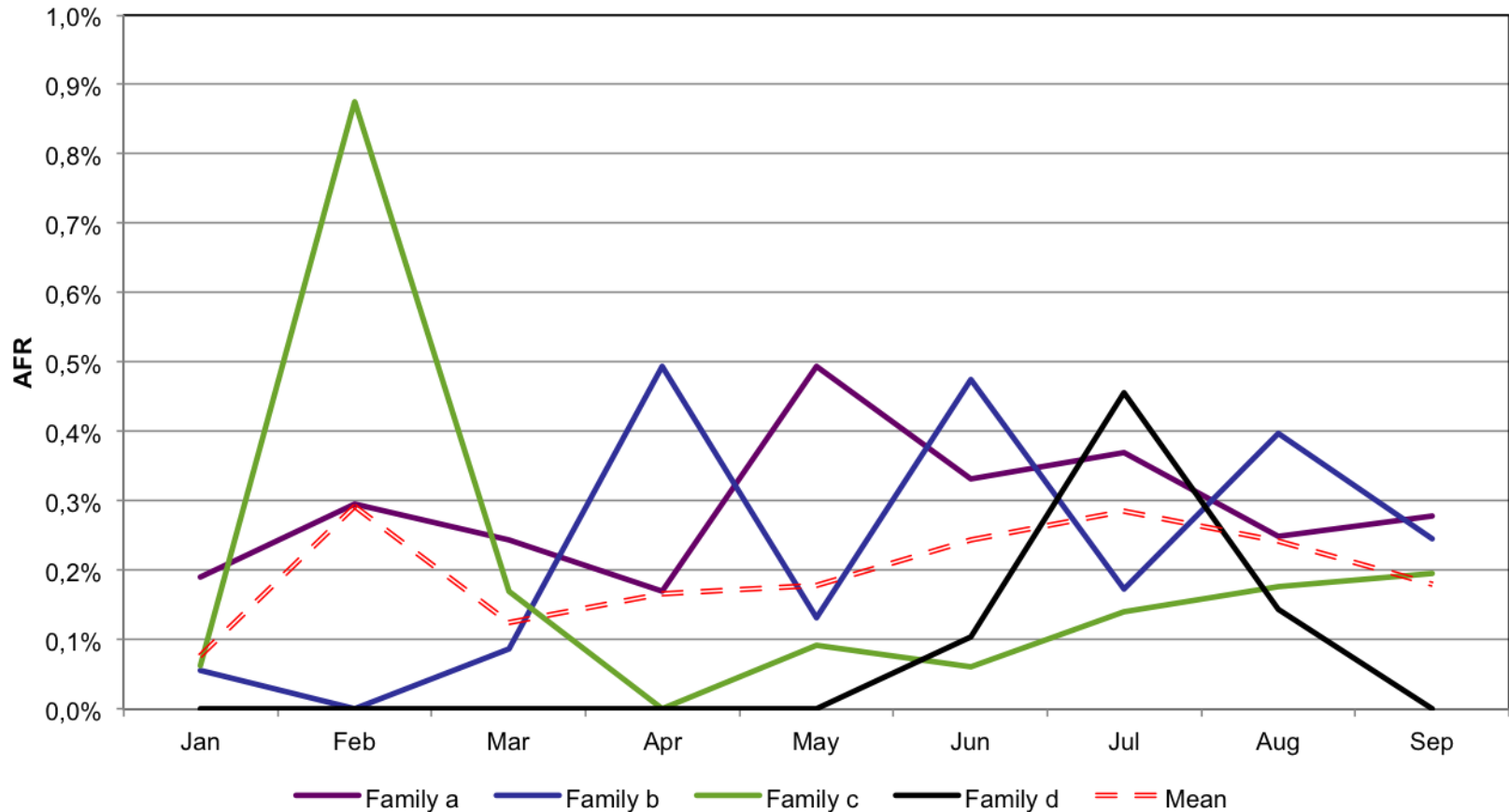


Full Factory Integration – Pre-cabled & pre-installed



Updated Numbers for Xyratex (2012)

Latest 4 families of SAS drives with volume shipment monthly field AFR



Actual AFR Data Experienced by Xyratex Sourced SAS Drives

Benefits of ClusterStor 6000

- ClusterStor 6000 is a highly available integrated Lustre storage solution providing end to end system management without sacrificing performance, leveraging industry standard components
- Built and optimized for Lustre
- High availability and redundancy is the basis of ClusterStor
- ClusterStor benefits from Xyratex strong roots in Disk Test equipment and traditional OEM solutions

Thank You

John_Fragalla@xyratex.com

xyratex